24.0 Digital Signal Processing

Academic and Research Staff

Prof. A.V. Oppenheim, Prof. J.S. Lim, Prof. B.R. Musicus, Prof. A.B. Baggeroer, G. Aliberti

Graduate Students

J. Bondaryk, D. Cobra, M. Covell, W.P. Dove, M. Feder, D. Griffin, D. Harasty, J. Hardwick, D. Izraelevitz, J. Jachner, T. Joo, D. Martinez, E. Milios, C. Myers, T. Pappas, J. Rodriguez, A. Silva, P. van Hove, M. Wengrovitz, A. Zakhor

Support Staff

P. Eiro, D. Gage

Part-time Assistants/Special Projects

S. Corteselli, A. Donato, M. Dove

Introduction

The Digital Signal Processing Group is carrying out research in the general area of signal processing. In addition to specific projects handled on campus, there is close interaction with Lincoln Laboratory and the Woods Hole Oceanographic Institution. While a major part of our activities focus on the development of new algorithms, there is a strong conviction that theoretical developments must be closely tied to applications. We are involved with the application areas of speech, image, video, and geophysical signal processing. We also believe that algorithm development should be closely tied to issues of implementation because the efficiency of an algorithm depends not only on how many operations it requires, but also on how suitable it is for the computer architecture it runs on. Also strongly affecting our research directions is the sense that while, historically, signal processing has principally emphasized numerical techniques, it will increasingly exploit a combination of numerical and symbolic processing, a direction that we refer to as knowledge-based signal processing.

In the area of knowledge-based signal processing, there are currently two research projects. One involves the concept of symbolic correlation, which is concerned with the problem of signal matching using multiple levels of description. This idea is being investigated in the context of vector coding of speech signals. Symbolic correlation will entail the use of both symbolic and numeric information to efficiently match a speech signal with stored code vectors. The second project in this area deals with the representation and manipulation of knowledge and expressions in the context of signal processing. This work examines issues such as the representation of knowledge, derivation of new knowledge from that which is given, and strategies for controlling the use of this knowledge.

In the area of speech processing, we have, over the past several years, worked on the development of systems for bandwidth compression of speech, parametric speech modeling, time-scale modification of speech and enhancement of degraded speech. Recently, a new model-based speech analysis/synthesis system has been developed. This system has been shown to be capable of high quality speech production, and it is currently being used in several mid-rate speech coding systems. Future work is aimed at using the speech model at lower bit rates through efficient coding techniques. Research is also continuing on adaptive noise cancellation techniques in a multiple microphone environment.

In image processing, we are pursuing a number of projects on restoration and enhancement. One project involves the estimation of coronary artery boundaries in angiograms. This research has produced a more robust model of the coronary angiograms, which consequently, improves the estimates of the arterial dimensions. A second image processing project is studying the removal of ghosts from television signals. This form of degradation is caused by multi-path channels and can be removed by the use of an appropriate inverse filter. The stable filter which results in general non-causal and, therefore, some form of time reversal must be used to implement the filter.

In the area of geophysical signal processing, current research is focused on the transformation of side scan sonar data. In practice, this data is corrupted by a number of factors related to the underwater environment. In this project, the goal is to explore digital signal processing techniques for extracting the topographic information from the actual sonographs. Concepts under study include the removal of distortions caused by towfish instability and reconstruction based on multiple sonographs taken from different angles.

There are also a number of projects directed toward the development of new algorithms with broad potential applications. For some time we have had considerable interest in the broad question of signal reconstruction from partial information, such as Fourier transform phase or magnitude. We have shown theoretically how, under very mild conditions, signals can be reconstructed from Fourier transform phase information alone. This work has also been extended to the reconstruction of multi-dimensional signals from one bit of phase, and, exploiting duality, zero-crossing and threshold crossing information. Current research in this area is aimed at reconstruction from distorted zero-crossings. In addition, the reconstruction from multiple threshold crossings is being studied. This problem has been shown to be better conditioned than reconstruction using only a single crossing. We are also examining the problem of narrowband signal detection in wideband noise. This project looks to compare several different techniques under a number of computational constraints. Research also continues on relationships between information theory and stochastic estimation. We are exploring applications to statistical problems, iterative signal reconstruction, short-time analysis/synthesis, and parameter estimation.

With the advent of VLSI technology, it is now possible to build customized computer systems of astonishing complexity for very low cost. Exploiting this capability, however, requires designing algorithms which not only use few operations, but also have a high degree of regularity and parallelism, or can be easily pipelined. The directions we are exploring include systematic methods for designing multi-processor arrays for signal processing, isolating signal processing primitives for hardware implementation, and searching for algorithms for multi-dimensional processing that exhibit a high degree of parallelism. We are also investigating highly parallel computer architectures for signal understanding, in which a mixture of intensive computation and symbolic reasoning must be executed in an integrated environment.

24.1 Reconstruction of Two-Dimensional Signals From Distorted Zero Crossings

National Science Foundation (Grant ECS 84-07285)
U.S. Navy - Office of Naval Research (Contract N00014-81-K-0742)

Alan V. Oppenheim, Joseph E. Bondaryk

It has been shown in theory that most two-dimensional, periodic, bandlimited signals are uniquely specified by their zero crossings or threshold crossings. It has also been proven that such two-dimensional images can in practice be reconstructed from threshold crossing information only.

It is the aim of this research to show that such an image, that has been affected by a non-linear, perhaps non-monotonic, distortion, can be reconstructed from the distorted signal's threshold crossing information. The threshold crossings will describe a set of linear equations, which can be solved to find the Fourier coefficients of the original image. An inverse DFT is then used to recover this image. The distortion may then be characterized by comparison of the original and distorted images. This research will attempt to identify the factors which most affect the reconstructability of the signal.

24.2 Digital Processing of Side Scan Sonar Data

National Science Foundation (Grant ECS 84-07285)
U.S. Navy - Office of Naval Research (Contract N00014-81-K-0742)

Alan V. Oppenheim, Daniel T. Cobra

Since its introduction in the early sixties, side scan sonar has proved to be a very important tool for underwater exploration, and in particular for marine geology. Several of its applications include surveying the sea floor, the search and location of objects on the bottom of the sea, and the prospection of mineral deposits.

The information contained in reflected sound waves is used by side scan sonar to produce a graphic record, called a sonograph, which constitutes a composite representation of the topographic features and the relative reflectivity of the various materials of the seabed. Due to several factors, however, sonographs do not provide a precise depiction of the topology. Geometric distortions can be caused by motion instability of the towfish on which the transducers are mounted, which can be caused by variable ship speeds and sea currents. The record can also suffer from interferences such as those caused by dense particle suspension in the water, shoals of fish, or by ultrasonic waves generated by passing ships. As a result, the extraction of topographic information from sonographs requires extensive practice and is often a tedious and time-consuming task.

Our general goal is to explore the application of digital signal processing techniques to side scan sonar data. At present, two specific problems are being contemplated. One is the estimation and correction of the distortions caused by towfish instability. The other is to determine the topography of the sea bed from the stereoscopic information contained in two or more sonographs of the same area made from different angles. The research is being conducted under M.I.T.'s joint program with the Woods Hole Oceanographic Institution, with the cooperation of the U.S. Geological Survey.

24.3 Representation and Manipulation of Signal Processing Knowledge and Expressions

National Science Foundation Fellowship
National Science Foundation (Grant ECS 84-07285)
U.S. Navy - Office of Naval Research (Contract N00014-81-K-0742)

Alan V. Oppenheim, Michele Covell

The phrase "signal processing" is used to refer to both "symbolic" and "numeric" manipulation of signals. "Symbolic" signal processing manipulates the signal description as opposed to the signal values with which "numeric" signal processing is primarily concerned. Efforts have been made to create computer environments for both types of signal processing.[1,2] Some issues that arise as a result of this work concern uniform representation of knowledge, derivation of new knowledge from that which is given, and strategies for controlling the use of this knowledge. This research will be concerned with these areas and how they apply to digital signal processing.

Representations that have been used in symbolic signal processing[1,2,3] have been largely distinct from those used in numeric signal processing.[1,4] The types of representations used are further separated by the control structures that the numeric and symbolic information commonly assume, the distinction essentially being the same as the distinction between Algol-like languages and logic programming languages. This dichotomy results from the differing amounts of available knowledge about the appropriate approaches to the problems being addressed. By separating the control structure from application knowledge, this dichotomy can be avoided.

Strategies for controlling when knowledge about a signal is used should be provided and new strategies should be definable, since these control structures provide additional information about the problem space, namely approaches that are expected to be profitable. Control strategies can also be used to outline new approaches to a problem, approaches that would not be considered by simple trigger-activated reasoning.

Finally, the ability to derive new knowledge from that which is given is desirable. This ability would allow the amount of information initially provided by the user to be minimized. The environment could increase its data base with new conclusions and their sufficient preconditions. Two immediate advantages of providing the environment with this ability are the reduction in the programming requirements and the possible "personalization" of the data-base. A reduction in programming requirements is available since information that is derivable from given information need not be explicity encoded. Commonly, this type of information is provided to improve the performance of the derivation process. Secondly, since the environment would add information to the data set according to conclusions prompted by the user's queries, the data set would expand those areas which the user had actively explored.

References

[1] Dove, W.P., "Knowledge-Based Pitch Detection," Ph.D. Diss., M.I.T., Cambridge, Mass., 1986.

[2] Kopec, G., "The Representation of Discrete-Time Signals and Systems in Programs," Ph.D. Diss., M.I.T., Cambridge, Mass., 1980.

[3] Milios, E., "Signal Processing and Interpretation using Multilevel Signal Abstractions," Ph.D. Diss., M.I.T., Cambridge, Mass., 1986.

[4] Myers, C., "Signal Representation for Symbolic and Numerical Processing," Ph.D. Diss., M.I.T., Cambridge, Mass., 1986.

24.4 Iterative Algorithms for Parameter Estimation with Applications to Array Processing

National Science Foundation (Grant ECS 84-07285)
U.S. Navy - Office of Naval Research (Contract N00014-81-K-0742)

Alan V. Oppenheim, Meir Feder, Ehud Weinstein[3]

Many signal processing problems may be posed as statistical parameter estimation problems. A desired solution for the statistical problem is obtained by maximizing the Likelihood (ML), the A-Posteriori probability (MAP) or some other criterion, depending on the a-priori knowledge. However, in many practical situations, the original signal processing problem may generate a complicated optimization problem, e.g., when the observed signals are noisy and "incomplete."

An iterative framework for maximizing the likelihood, the EM algorithm, is widely used in statistics. In the EM algorithm, the observations are considered "incomplete" and the algorithm iterates between estimating the sufficient statistics of the "complete data" given the observations and a current estimate of the parameters (the E step), and maximizing the likelihood of the complete data, using the estimated sufficienct statistics (the M step). When this algorithm is applied to signal processing problems, it yields, in many cases, an intuitive appealing processing scheme.

In the first part of this research, we investigate and extend the EM framework. By changing the "complete data" in each step of the algorithm, we can achieve algorithms with better convergence properties. In addition, we suggest EM type algorithms to optimize other (non ML criteria). We also develop sequential and adaptive versions of the EM algorithm.

---

[3] Woods Hole Oceanographic Institution

In the second part of this research, we examine some applications of this extended framework of algorithms. In particular, we consider:

- Parameter estimation of composite signals, i.e., signals that can be represented as a decomposition of simpler signals. Examples include:

    1) multiple source location (or bearing) estimation and

    2) multipath or multi-echo time delay estimation.

- Noise cancellation in a multiple microphone environment (speech enhancement)

- Signal reconstruction from partial information (e.g., Fourier transform magnitude).

The EM-type algorithms suggested for solving the above "real" problems provide new and promising procedures, and they thus establish the EM framework as an important tool to be used by a signal processing algorithm designer.

24.5 A New Mixed-Excitation Speech Model and Its Application to Bit-Rate Reduction

National Science Foundation (Grant ECS 84-07285)
Sanders Associates, Inc.
U.S. Air Force - Office of Scientific Research (Contract F19628-85-K-0028)
U.S. Navy - Office of Naval Research (Contract N00014-81-K-0742)

Jae S. Lim, Daniel W. Griffin

One approach to speech coding involves estimating and encoding the parameters associated with some underlying speech model (vocoders). The encoded speech model parameters are then transmitted and the receiver decodes them, employing the speech model to synthesize speech. Examples of vocoders include linear prediction vocoders, homomorphic vocoders, and channel vocoders. In these vocoders, speech is modeled on a short-time basis as the response of a linear system excited by a periodic impulse train for voiced sounds or random noise for unvoiced sounds. For this class of vocoders, speech is analyzed by first segmenting speech using a window such as the Hamming window. Then, for each segment of speech, the excitation parameters and system parameters are determined. The excitation parameters consist of the voiced/unvoiced decision and the pitch period. The system parameters are used to synthesize an excitation signal consisting of a periodic impulse train in voiced regions and random noise in unvoiced regions. This excitation signal is then filtered using the estimated system parameters.

Even though vocoders based on this underlying speech model have been quite successful in synthesizing intelligible speech, they have not been successful in synthesizing high quality speech. As a consequence, they have not been widely used in mid-rate coding applications requiring high quality speech reproduction. The poor quality of the synthesized speech is due to fundamental limitations in the speech models, inaccurate estimation of the speech model parameters, and distortion caused by the parameter coding.

The goal of this research is to develop an improved speech model, together with methods to estimate the speech model parameters robustly and accurately, and apply the resulting speech analysis/synthesis system to the problem of bit-rate reduction. In this model, the short-time spectrum of speech is modeled as the product of an excitation spectrum and a spectral envelope. The spectral envelope is some smoothed version of the speech spectrum. The excitation spectrum is represented by a fundamental frequency, a voiced/unvoiced (U/UV) decision for each harmonic of the fundamental, and the phase of each harmonic declared voiced. In speech analysis, the model parameters are estimated by explicit comparison between the original speech spectrum and the synthetic speech spectrum. Current work is concentrated on efficient coding of these speech model parameters. Preliminary results indicate that very good quality reproduction can be obtained with this speech coding system for both clean and noisy speech without the "buzziness" and severe degradation (in the presence of noise) typically associated with vocoder speech rate at a rate of 9.6 kbps.

24.6 Mixed Causuality IIR Filtering for Nonminimum Phase Channel Compensation

National Science Foundation (Grant ECS 84-07285)
U.S. Navy - Office of Naval Research (Contract N00014-81-K-0742)

Alan V. Oppenheim, Daniel J. Harasty

A feedforward (or multipath) channel models the physical situation which gives rise to television ghosts. Digital compensation for these artifacts can be achieved by designing a filter which is the inverse of the ghosting filter. This filter, however, is stable and causal only if the ghosting filter is minimum phase, that is, all of its singularities lie inside the unit circle. If the ghoster is nonminimum phase, the stable inverse filter has an infinite impulse response which extends backwards in time. Such an impulse response can be achieved if the data can be run backwards through a regular IIR filter. Since the prospect of having the entire video signal in storage to do the filtering is unlikely, investigation of a method of piece-wise time reversal is necessary.

Basically, the MCIIR system is approximated by separating the deghosting system into two cascadable systems: the first has a strictly causal impulse response corresponding to the poles inside the unit circle, and the second has a strictly anticausal impulse response corresponding to the poles outside the unit circle. The ghosted signal is fed normally through the causal system, and then broken into blocks to be time flipped, fed through the time-reversed implementation of the anticausal portion, and flipped again upon output. Unfortunately, the anticausal portion is not in the correct state when a new block is introduced, and so a portion of the outputed block (of a length equal to the effective length of the anticausal filter) must be disregarded as corrupted by the transient response of the filter. Consequently, overlapping blocks must be used to obtain valid values for all output samples.

24.7 A 4.8 Kbps High Quality Speech Coding System

National Science Foundation (Grant ECS 84-07285)
U.S. Navy-Office of Naval Research (Contract N00014-81-K-0742)

Jae S. Lim, John C. Hardwick

Recently completed research has led to the development of a new speech model. This model has been shown to be capable of producing speech without the artifacts common to model based speech systems.[1] This ability makes the model particularly applicable to speech coding systems requiring high quality reproduction at a low bit rate. A 9.6 kbps based on this model is documented.[2] In addition, slight modifications of this system have been made which extend it to 8 kbps. Both of these systems have been shown to be capable of high quality output.

The purpose of this research is to explore methods of using the new speech model in a 4.8 kbps speech coding system. Results indicate that a substantial amount of redundancy exists between the model parameters. Current research efforts are aimed at methods of exploiting this redundancy in correspondence with more efficient quantization of the model parameters. Preliminary experiments verify that these techniques will permit a substantial bit reduction without major degradations in speech quality.

References

[1] Griffin, D.W., and J.S. Lim, "A New Model-Based Speech Analysis/Synthesis System," I.E.E.E. International Conference on Acoustics, Speech and Signal Processing , Tampa, Florida, 1985, pp. 513-516.

[2] Griffin, D.W., and J.S. Lim, "A High Quality 9.6 kbps Speech Coding System," I.E.E.E. International Conference on Acoustics, Speech and Signal Processing , Tokyo, Japan, 1986.

24.8 Multi-Representational Signal Matching

Canada, Bell Northern Research Scholarship
Canada, Fonds pour la Formation de Chercheurs et l'Aide a la Recherche Postgraduate Fellowship
National Science Foundation (Grant ECS 84-07285)
Canada, Natural Science and Engineering Research Council Postgraduate Fellowship
U.S. Navy - Office of Naval Research (Contract N00014-81-K-0742)

Alan V. Oppenheim, Jacek Jachner

This work investigates the idea of performing signal matching by integrating different signal representations in the matching procedure. In as much as the signal representations can be numeric or symbolic, multi-representational matching is an approach to signal matching in the case when the signals are described by a mix of numeric and symbolic representations.

As a framework in which to study signal matching, current work is focusing on the vector coding of speech signals. Recent progress in speech waveform coders has demonstrated the feasibility of high quality speech at very low bit rates (less than 8 kb/s) using an adaptive predictive coder strategy with block quantization of the prediction residual. The computations required to select an optimum vector codeword, even for moderately sized codebooks, are prohibitively long. Using an implementation of such a coder in the KBSP package on LISP machines, our current focus is to efficiently match a given sequence of speech data with a stored codebook of vectors by using multiple signal representations.

24.9 Detection of Narrowband Signals in Wideband Noise

National Science Foundation (Grant ECS 84-07285)
Sanders Associates, Inc.
U.S. Navy - Office of Naval Research (Contract N00014-81-K-0472)

Alan V. Oppenheim, Tae H. Joo

The search for radio signals transmitted by extraterrestrials is a complex, multidimensional search problem because little is known about the transmitted signal. Current searches for extraterrestrial intelligence (SETI) collect data from a predetermined range of signals. These data are then processed to detect all synthetic components. (Synthetic components of signals are those which do not originate naturally. This assumes that the synthetic component is generated by extraterrestrials.) The assumption that the transmitted signal is a continuous wave (CW) at certain frequencies is commonly used in determining the range. Existing SETI systems use a frequency of 1450 MHz, atomic hydrogen line.

Due to uncertainties in the transmitter locations, the relative velocities and the receiver antenna beamwidth, the frequency of the CW signal is unknown but is within 200 KHz of 1420 MHz. The propagation experiences multi-path which spreads the CW signal to a bandwidth of about 0.05 Hz. Therefore, SETI systems must search a wide frequency band (approximately 400 KHz) to detect a very narrowband (0.05 Hz) signal in poor signal-to-noise ratio (SNR) conditions.

Current SETI systems use FFT's to compute the spectrum. Each spectrum is then compared to a threshold to detect a peak. Because the SNR is low, the frequency bin size of the FFT is matched to the bandwidth of the narrowband signal. Therefore a 2^{23}, or approximately 400 KHz/0.05 Hz, length FFT is required. In an existing system known as mega-channel extraterrestial array (META)[1], this FFT is computed in two steps. First, the signal is filtered by 128 band-pass filters. Second, each band-pass filtered signal is transformed by a 64K length FFT. These computations are made using fixed point arithmetic. There are alternative implementations of this DFT-based method. The performance of different implementations, within constraints of the finite register length and other computational limitations, will be examined.

If the received signal is modelled as a sinusoid in white noise, modern spectrum estimators (e.g., the maximum entropy method) or frequency estimators (e.g., Pisarenko's method) can be employed. The performance and applicability of these algorithms, within constraints of computational limitations will be examined.

Reference

[1] Horowitz, P., J. Forster and I. Linscott, "The 8-Million Channel Narrowband Analyzer," in *The Search for Extraterrestrial Life: Recent Developments*, edited by M.D. Papagiannis. Hingham, Mass.: Kluwer Academic Publishers, 1985, pp. 361-371.

24.10 Estimation of Coronary Artery Boundaries in Angiograms

National Science Foundation (Grant ECS 84-07285)
U.S. Navy - Office of Naval Research (Contract N00014-81-K-0742)

Jae S. Lim, Thrasyvoulos N. Pappas

The precise and objective measurment of the severity of coronary obstructions from coronary angiograms is important in the treatment of patients with ischemic heart disease. An angiogram is an x-ray picture of the coronary arteries in which a contrast agent has been injected via a catheter.

Coronary artery imaging presents special problems because of the arteries' location on the beating heart and their shape and size. As a result, many techniques that have been used quite successfully for stenosis determination of other arteries, like the femoral and carotid, do not have satisfactory performance when applied to coronary arteries. These algorithms are quite heuristic and find the boundaries of the artery as the inflection points of a series of densitometric profiles perpendicular to the vessel image. Even though this approach is computationally simple, there is no theoretical justification for it.

We consider a different approach which more fully exploits the detailed characteristics of the signals involved. Specifically, we develop a model of the film density of the coronary angiograms and use it to estimate the diameter and cross-sectional area at each point along the vessel. Our model accounts for the structure of the vessel and background, as well as the distortions introduced by the imaging system (blurring and noise). We have developed both a one-dimensional model of the density profiles perpendicular to the vessel image, and a two-dimensional model of rectangular sections of the image. The parameters of the model include the vessel center-point locations and the radius at each point. The spatial continuity of the vessel is incorporated into the model and contributes to the accuracy of the estimation procedure. The algorithms are tested on synthetic data, on x-rays of contrast-medium-filled cylindrical phantoms obtained over a wide range of radiographic conditions, and on real coronary angiograms. Our results indicate that the 1-D algorithms have better performance than current methods, and preliminary results indicate that the 2-D algorithms have better performance than 1-D algorithms.

24.11 Reconstruction of Multidimensional Signals from Multilevel Threshold Crossings

Fanny and John Hertz Foundation Fellowship
National Science Foundation (Grant ECS 84-07285)
U.S. Navy - Office of Naval Research (Contract N00014-81-K-0742)

Alan V. Oppenheim, Avideh Zakhor

It has been shown theoretically that under mild conditions multidimensional signals can be recovered from one bit of phase. Exploiting duality, they can also be recovered from zero crossings or one-level crossings. In practice, recovery from zero crossings, which is essentially an implicit sampling strategy, is highly ill-considered. That is, the number of position bits of zero crossings required for reconstruction is rather large. One way to improve robustness is to reconstruct from multilevel threshold crossings as opposed to one level.

In this research, we extend some of the well-known results in interpolation theory in order to define a new implicit sampling strategy which can ultimately be used to find the sufficient condition for recovery from multilevel threshold crossings. We then use algebraic geometry ideas to find conditions under which a signal is almost always reconstructable from its multilevel threshold crossings. As it turns out, the set of signals which do not satisfy this latter condition are of measure zero among the set of all bandlimited, periodic two-dimensional signals.

Simulations are used to verify the above ideas experimentally. It is shown that the reconstruction becomes more robust as the number of threshold levels are increased. Furthermore, the robustness is improved if more points are used for reconstruction. At the extreme case, where all the points corresponding to each threshold are used, the theory of projections of convex sets can be used to derive an iterative algorithm for reconstruction.

Finally, the ideas developed for recovery from level crossings will be applied to the problem of reconstruction of continuous images from half-tone ones.

24.12 Knowledge-Based Pitch Detection

National Science Foundation (Grant ECS 84-07285)
Sanders Associates, Inc.
U.S. Navy - Office of Naval Research (Contract N00014-81-K-0742)

Alan V. Oppenheim, Webster P. Dove, Randall Davis

Knowledge-based signal processing (KBSP) is the application of the numerical algorithms of signal processing and the symbolic inference techniques of expert systems on problems where both numerical and symbolic information are present. One class of problems in this area includes those for which both numeric and symbolic input information is available. The Knowledge-Based Pitch Detection (KBPD) project prepared a system to describe the excitation of a sentence of speech (voicing and periodicity) given a recorded waveform and a phonetic transcript of the sentence. Through this project we learned more about the general problem of integrating numeric and symbolic processing.

This problem is interesting because pitch detection is an old problem with no completely successful solution (for high quality resynthesis). Existing programs attack the problem with a single unified algorithm based on the detection of periodicity. Typically, this will work some places and not others. For example, it is common for voiced sections at the ends of sentences to be irregular and not periodic. Algorithms based on periodicity cannot conclude voicing in such regions and therefore they make voiced-to-unvoiced errors there.

With more information available to it, the KBPD system can avoid such problems. Since it has a transcript, KBPD starts with strong information about the voicing of the sentence in the middle of the phonemes of the transcript. However, aligning the transcript precisely with the sentence so that the boundaries of the phonemes are correct require interaction between the transcript representation and numerical measurements of the waveform. Also, the actual rendition of a given phoneme may vary enough between speakers and sentences that a normally voiced phoneme is not actually voiced. So the system must verify the transcript against the observed properties of the sentence to avoid mistakes.

By combining the results of several different methods of analysis rather than relying on a single algorithm, KBPD spreads the risk of errors. To the extent that these different methods both cooperate effectively and span the phenomena of the problem, one can expect the system to be more robust than a single algorithm approach. The key to this cooperation is the explicit representation of credibility in the results of each component.

Expert systems have capitalized on the representation of credibility for performing symbolic inference when both knowledge and problem information are uncertain. KBPD moves these techniques to the domain of numeric processing with the explicit representation in the program of both certainty and accuracy for numeric results. The extra information (beyond the basic result values) allows the system to merge results from different components without specific information about the nature of the components. This in turn allows the integration of new components into the system without major rewriting of existing code, and allows the final results to be based on the components that rated themselves most credible. While there is some question as to how reliably a given algorithm can be expected to rate its own performance, clearly this is an improvement over the conventional program assumptions that the algorithm's result "are the answer," and it has the virtue of presenting to the operator some information about the overall credibility of results.

This project also developed a program called the Pitch Detector's Assistant (PDA). It serves as both an assistant in the pitch detection process, and a workbench to evaluate numeric and symbolic processing techniques in pitch detection programming.

As an end product, it might be used to generate excitation information for off-line speech storage applications (like talking appliances) or for speech research into the properties of pitch as an information carrying medium. It is clearly not for real-time applications like vocoding, though it might be used for vocoding experiments.

As a workbench for pitch detection programming, the PDA allows us to assess what kind of symbolic information is useful in the pitch detection problem by experimenting with ways of using it.

As a KBSP catalyst, it has prompted the development of new data structures (for representing temporal estimates and pitch estimates). Furthermore, it has motivated the development of a signal processing environment (called KBSP) that dramatically reduces the time to implement signal processing algorithms. This software environment is also capable of embedding symbolic processing functions (like the generation of symbolic assertions) within numerical algorithms. That capacity, together with a rule-based system capable of invoking signal processing functions, permits a flexible, two-way communication between the numeric and symbolic parts of the system. Finally, this project has led to the clarification of some of the features of symbolic inference in the context of numeric information. Specifically, we have found that recognition of equivalent assertions is more difficult in a numeric context, and that there are more ways of combining numeric assertions than symbolic ones.

This project was completed in June 1986.

24.13 Reconstruction of a Two-Dimensional Signal from the Fourier Transform Magnitude

National Science Foundation (Grant ECS 84-07285)
U.S. Navy - Office of Naval Research (Contract N00014-81-K-0742)

Jae S. Lim, David Izraelevitz

In the case of an arbitrary multidimensional discrete signal, the phase and magnitude of the Fourier transform are independent functions of frequency. However, in many situations there is additional information regarding the signal which provides a very strong connection between the phase and magnitude. Specifically, it has been shown that almost all multidimensional signals which are non-zero only over a specified domain are uniquely specified, in a sense, by knowledge of the Fourier transform magnitude alone. Several algorithms have been developed for reconstructing such a signal from its Fourier transform magnitude; however, they all fall into either of two categories: they are heuristic algorithms which sometimes do not converge to the true reconstruction, or they are computationally too expensive for even moderate size signals.

In this research, we present a new algorithm for reconstruction from Fourier transform magnitude which is a closed form solution to the problem and which has been used to reliably construct signals of extent up to 20 by 20 pixels. The algorithm is based on posing the problem of Fourier transform magnitude reconstruction as requiring the explicit factorization of the z-transform of the autocorrelation sequence of the unknown function. This z-transform is easily computed from knowledge of the Fourier transform magnitude. A new procedure is developed for factoring large bivariate polynomials and this algorithm is applied to the needed factorization of the autocorrelation z-transform.

This project involves studying the data noise sensitivity of the algorithm and of the problem of reconstruction from Fourier transform magnitude in general, as well as developing a detailed comparison of the behavior of the present algorithm with previously proposed algorithms.

This project was completed in May 1986.

24.14 Model-Based Motion Estimation and its Application to Restoration and Interpolation of Motion Pictures

Center for Advanced Television Studies
National Science Foundation (Grant ECS 84-07285)
U.S. Navy - Office of Naval Research (Contract N00014-81-K-0742)

Jae S. Lim, Dennis M. Martinez

Motion pictures can be manipulated in a variety of ways of compensating for motion within the image sequence. An important component of all motion-compensated image processing systems is motion estimation. Previous approaches to motion estimation have encountered two primary problems, computational complexity and estimation accuracy. This work is concerned with the development and analysis of computationally efficient motion estimation algorithms which can determine motion trajectories very accurately.

A model-based motion estimation algorithm has been developed. This algorithm requires significantly less computation than traditional approaches. In addition, it can determine velocity fields more accurately than commonly used region matching methods. The algorithm is based on a local three-dimensional signal model and a local translational velocity model. It is possible to estimate complex velocity fields encountered in real-life television images with the algorithm.

We applied this algorithm to several problems in motion picture restoration and interpolation. Image restoration algorithms which operate on a single frame at a time usually remove noise at the expense of picture sharpness. However, we demonstrated that motion-compensated restoration systems can remove noise with little or no loss in picture sharpness.

The algorithm has also been used successfully for frame interpolation. A motion-compensated frame interpolation system was developed which permits computing frames at arbitrary times. This system can be used in a variety of applications involving frame rate modification. A number of experiments have shown that this motion-compensated interpolation system produces motion pictures with better motion rendition than traditional frame repetition systems.

This project was completed in September 1986.

24.15 Signal Processing and Interpretation using Multilevel Signal Abstractions

National Science Foundation (Grant ECS 84-07285)
Sanders Associates, Inc.
U.S. Navy - Office of Naval Research (Contract N00014-81-K-0742)

Alan V. Oppenheim, Evangelos E. Milios

A signal processing system which integrates signal and symbol processing was developed for acoustic waveform processing and interpretation. The design of the system was derived from the systematic observation (protocol collection) and subsequent analysis of human signal processing activity. The resulting system consists of: 1) a harmonic-set formation subsystem, which produces harmonic sets present in acoustic spectra and their symbolic description; and 2) a geometrical hypothesis formation and testing subsystem, which forms hypotheses about the acoustic source motion based on the data, and then performs detailed testing against the data. The system is being built using a hybrid methodology combining both procedural and declarative programming and accommodates both algorithmic and heuristic techniques in signal processing. Modules perform spectral peak analysis using rules that stem from human perceptual considerations. The rule-based approach is enhanced to allow thresholds to be set from examples through the same rules used for classification. In comparison to previous signal/symbol processing systems, which relied mostly on symbol processing, this system is based on the concept that a tight interaction of signal processing and interpretation can save a lot of symbol processing.

This work was done in collaboration with the Machine Intelligence Technology Group at the M.I.T. Lincoln Laboratory. This project was completed in June 1986.

24.16 Signal Representation for Symbolic and Numerical Processing

Amoco Foundation Fellowship
National Science Foundation (Grant ECS 84-07285)
Sanders Associates, Inc.
U.S. Navy - Office of Naval Research (Contract N00014-81-K-0742)

Alan V. Oppenheim, Cory S. Myers

One important aspect of our work on Knowledge-Based Signal Processing (KBSP) is the development of a suitable environment for carrying out the research and for exploring mechanisms for integrating numerical and symbolic processing. As part of our KBSP work, we have developed the KBSP software package, an innovative software package for signal processing on the LISP Machine. The KBSP software package provides a uniform signal processing package that is integrated into an interactive, symbolic processing environment, the LISP Machine. The package provides computational speed, ease of development, and a close correspondence between abstract signal processing concepts and programming implementation.

As an outgrowth of the KBSP package, this research is devoted to the incorporation of symbolic manipulation facilities into the numerical KBSP spackage. Symbolic manipulation of signals involve the manipulation of signal representations rather than signal values. One example of this symbolic manipulation is the symbolic generation of Fourier transforms. The system understands many primitive signals, their Fourier transforms, and rules for the manipulation of Fourier transforms with respect to other signal processing operations, such as addition, multiplication, and convolution. For many signals, the system is able to parse the signal's representation and generate its Fourier transform without any reference to the numerical values of the signal.

One area in which the symbolic manipulation of signal representation is a natural one is the area of varying the signal processing according to "context". Context can include properties of the signals under consideration, properties of the hardware, or properties of the world. For example, different FFTs can be used when the signal is real-valued or when it is complex-valued, different filtering algorithms can be used to achieve different trade-offs among multipliers, adds, and memory references, and different spectral modeling techniques can be used for different speech sounds. The goal of this research is to automate the decision process used in algorithm selection.

As an initial step in automating the algorithm decision process, we have developed a system for the LPC analysis of a speech signal given a time-aligned phonetic transcript of the speech signal and a symbolic description of the recording environment. In this system, the parameters of the LPC analysis -- the number of poles, the window size, and the frame rate -- are varied according to the information present in the phonetic transcript. For example, a long window and a low frame rate are used in the steady-state portion of vowels and a short window and a high frame rate are used in the area of stop releases.

We are currently working on more general techniques that will allow many different sources of information to be used in choosing a signal processing algorithm. One aspect of this problem is the specification of properties of both signals and systems. We are trying to develop a symbolic signal processing language in which standard signal processing properties such as finite-durationness, symmetry, linearity, shift-variance, etc., are easily represented and manipulated by the system.

We are also studying methods for the manipulation of signal representations so that the system can determine equivalent forms for algorithm implementation. For example, the system will be able to implement a filter in either the time-domain or the frequency-domain. This capability will be used to automatically choose different algorithms to implement the same signal processing operation for different input signals and for different trade-offs among the different factors that affect computational cost.

This project was completed in August 1986.

24.17 Reconstruction of Undersampled Periodic Signals

National Science Foundation (Grant ECS 84-07285)
U.S. Navy - Office of Naval Research (Contract N00014-81-K-0742)

Alan V. Oppenheim, Anthony J. Silva

Under certain circumstances, it may be impossible to sample a signal at a rate greater than twice the highest spectral component present. In particular, the sampling rate might be limited by the A/D converter and associated hardware used. However, for certain classes of undersampled signals, complete recovery is still possible (at least theoretically), in spite of the well-known Nyquist criterion.

Periodic signals form one such class. Consider undersampling a periodic signal whose harmonic frequencies are mutually prime to the sampling rate. The harmonics have zero bandwidth[4] and though aliased into the baseband (-Fsamp/2 to + Fsamp/2 Hz), they do not overlap. The Nyquist criterion is stated generally for low-pass waveforms with energy presumably spread smooth across their spectra. Because of this, it discounts the possibility of recovery after the non-destructive aliasing above.

In what follows, we assume the sampling rate is stable and known to a modest degree of accuracy. When the signal period is known, recovery is trivial. The time sample x[n] are sorted by interpreting the index "n" modulo the normalized signal period[5] then placing each sample in the appropriate place in a "composite" period. No samples will overlap if the mutual primality requirement above is met.

A far more interesting problem exists when the signal period is known. Rader[1] has presented an iterative recovery technique in which a multiplicity of trial signal periods is used as moduli in the sorting process above. The trial period yielding the best reconstruction is retained as the estimate of the true signal period. Results from number theory (Farey sequences, modular or congruential arithmetic, etc.) are exploited to make the approach practicable.[2]

We searched for ways to accelerate the Rader algorithm. An analogous frequency domain algorithm, also employing number theory, was developed. It consisted of determining the frequencies and amplitudes of the aliased harmonics, sorting them, and inverse transforming. Performance of all algorithms and their variants in the presence of noise, slightly wavering signal frequency and amplitude, and other undesirable conditions, were examined.

This project was completed in January 1986.

References

[1] C.M. Rader, "Recovery of Undersampled Periodic Waveforms," I.E.E.E. Trans. Acoust., Speech, Signal Process., *ASSP-25*, 3, 1977.

[2] G.H. Hardy and E.M. Wright, *An Introduction to the Theory of Numbers*, Oxford University Press, 1956.

---

[4] Of course, the signal would have to be known and sampled for all time to yield truly zero-bandwidth harmonics.

[5] The ratio of the signal and sampling periods.

24.18 Silhouette-Slice Theorems

National Science Foundation (Grant ECS 84-07285)
U.S. Navy - Office of Naval Research (Contract N00014-81-K-0742)

Jae S. Lim, Patrick van Hove

This research addresses the relationships between an object and the boundaries of its silhouettes, which are referred to as contours, corresponding to various three-dimensional (3-D) orientations of the line of sight. For this purpose, special models of objects and silhouettes are considered. The property sphere of an object is defined on a unit sphere which is related to the object by a 3-D Gaussian mapping. Similarly, the property circle is related to the contour it represents by a 2-D Gaussian mapping. In earlier computer vision work, property spheres and circles have been used independently, and their applications have been limited to the representation of scalar fields of object properties.

In a first stage, we have shown how the concepts of property spheres and circles can be usefully combined to relate the properties of an object and those of its contours. Specifically, it is proved that a slice through the property sphere of an object leads to the property circle of the contour corresponding to the line of sight perpendicular to that slice.

In a second stage, a new concept of object modeling has been introduced, where the property sphere is used as a domain for vector and tensor fields of object properties. In particular, a new representation of 3-D objects and 2-D silhouettes, referred to as the Curvature Transform (CT), maps the inverse of the curvature tensor field of the surface of an object on its property sphere, and the radius of curvature of the contour of a silhouette on its property circle. The key advantage of this representation is that a slice through the CT of an object followed by projection of the tensor field produces the CT of the contour corresponding to the line of sight perpendicular to the slice.

The study progressed with attempts to use these new concepts in the reconstruction of object shape from silhouette information. Surface reconstruction procedures have been proposed in the field of machine vision, in the contexts of shape from photometric stereo, shading, texture, and motion. These procedures reconstruct a viewer-dependent 2 1/2-D sketch of the surfaces. The surface reconstruction procedure which is attempted here would provide a full 3-D sketch of the objects.

This work was done in collaboration with the Distributed Sensor Systems group at the M.I.T. Lincoln Laboratory. This project was completed in September 1986.

24.19 The Hilbert-Hankel Transform and Its Application to Shallow Water Ocean Acoustics

Fanny and John Hertz Foundation Fellowship
National Science Foundation (Grant ECS 84-07285)
U.S. Navy - Office of Naval Research (Contract N00014-81-K-0742)

Alan V. Oppenheim, George V. Frisk, Michael S. Wengrovitz[6]

The problem of continuous-wave acoustic field propagation in shallow water is being investigated. In this environment, components of the field alternately reflect off both the ocean surface and the ocean bottom. In effect, the water can be considered as an acoustic waveguide, bounded by the ocean surface and the underlying ocean bottom. Several aspects of this waveguide propagation problem are being studied. The first concerns the development of an accurate numerical model to predict the magnitude and phase of the acoustic field as a function of range from the source, given the geoacoustic parameters of the water and the bottom. A technique has been developed which computes the field based on its decomposition into trapped (resonant) and continuum (non-resonant) components.

A second aspect being studied is the application of the Hilbert-Hankel transform to both the synthesis and inversion of these fields. This transform has a number of interesting and useful properties and forms the basis for a reconstruction method in which the real (imaginary) part of the complex-valued field is obtained from the imaginary (real) part.

Finally, the inherent sensitivity of extracting the bottom plane-wave reflection coefficient from measurements in the reverberant waveguide environment was researched. Results indicate that there are several invariant waveguide parameters which do not depend on the properties of the underlying media. By exploiting the invariance of these parameters, it is possible to design an actual ocean experiment from which an improved reflection coefficient estimate results.

This project was completed in January 1986.

24.20 Speech Enhancement Using Adaptive Noise Cancellation

National Science Foundation Fellowship
National Science Foundation (Grant ECS 84-07285)
U.S. Air Force - Office of Scientific Research (Contract F19628-85-K-0028)
U.S. Navy - Office of Naval Research (Contract N00014-81-K-0742)

Jae S. Lim, Jeffrey J. Rodriguez

In many military environments, such as fighter jet cockpits, the increasing use of digital communication systems has created a need for robust vocoders and speech recognition systems. However, the high level of acoustic noise makes the vocoders less intelligible and makes reliable speech recognition more difficult. Therefore, we are using Widrow's Adaptive Noise Cancelling (ANC) algorithm to enhance the noise-corrupted speech.[1]

ANC is a noise-reduction method that uses multiple inputs. In the fighter jet application, we use two microphones: a primary and a reference. The primary microphone

---

6 Woods Hole Oceanographic Institution

is located inside the pilot's oxygen face mask, and the reference microphone is attached to the exterior of the face mask. In this configuration, the primary microphone records the noise-corrupted speech, while the reference microphone ideally records only the noise. When these two inputs are processed, the reference noise is filtered and subtracted from the primary signal. Hopefully, the resulting signal will contain the undegraded speech signal with very little noise.

This project was completed in January 1986.

Reference

[1] B. Widrow, et al., "Adaptive Noise Cancelling: Principles and Applications," Proc. I.E.E.E., *63*, 12, (1975).